

HIGH DYNAMIC RANGE IMAGE TONE MAPPING BY MAXIMIZING A STRUCTURAL FIDELITY MEASURE

Hojatollah Yeganeh and Zhou Wang

Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Email: hyeganeh@uwaterloo.ca, zhouwang@ieee.org

ABSTRACT

Tone mapping operators (TMOs) that convert high dynamic range (HDR) images to standard low dynamic range (LDR) images are highly desirable for the visualization of these images on standard displays. Although many existing TMOs produce visually appealing images, it is until recently validated objective measures that can assess their quality have been proposed. Without such objective measures, the design of traditional TMOs can only be based on intuitive ideas, lacking clear goals for further improvement. In this paper, we propose a substantially different tone mapping approach, where instead of explicitly designing a new computational structure for TMO, we search in the space of images to find better quality images in terms of a recent objective measure that can assess the structural fidelity between two images of different dynamic ranges. Specifically, starting from any initial image, the proposed algorithm moves the image along the gradient ascent direction and stops until it converges to a maximal point. Our experiments show that the proposed algorithm reliably produces better quality images upon a number of state-of-the-art TMOs.

Index Terms— tone mapping, structural fidelity, high dynamic range image, image quality assessment, gradient-based optimization

1. INTRODUCTION

With the rapid advances of imaging and computer graphics technologies, high dynamic range (HDR) images that have greater dynamic ranges of intensity levels than standard or low dynamic range (LDR) images are becoming widely available [1]. In most application environment, in order to visualize such HDR images, tone mapping operators (TMOs) must be applied that convert them to LDR images before they can be presented using standard displays. A significant number of TMOs have been proposed recently (for example [2–5]), many of which are able to produce visually appealing pictures. However, these TMOs were developed without the question of “what constitutes a good tone mapped image” being properly answered. Quality assessment methods of tone mapped images are highly desirable, without which tone mapping research lacks solid scientific ground — algorithm development has no optimization goal, different TMOs cannot be compared, and further improvement is directionless.

Subjective evaluation is the most straightforward method to assess tone mapped images [6–8] but is extremely time consuming and expensive, making them impossible to be embedded into automatic optimization algorithms [9]. In addition, certain image structures contained in HDR images may be mapped to flat regions (and are thus missing) in tone mapped images, but human observers may not be aware of their existence. Therefore, subjective evaluation may

not be regarded as a fully reliable golden standard in assessing the quality of tone mapped images.

Objective quality assessment of tone mapped images is a difficult problem. Since the original HDR and the tone mapped LDR images have different dynamic ranges, typical objective image quality measures such as peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [10, 11] are not applicable. To the best of our knowledge, only a few objective assessment methods have been proposed for HDR images. The HDR visible differences predictor (HDR-VDP) [1, 12] attempts to differentiate between visible and invisible distortions in terms of detection probability. However, HDR-VDP is designed to predict the visibility of distortions between two HDR images of the same dynamic range, and cannot be employed to compare an HDR with an LDR image. A dynamic range independent quality measure was proposed in [13], which improves upon HDR-VDP and produces three quality maps indicating the loss of visible features, the amplification of invisible features, and the reversal of contrast polarity, respectively. Because of the reduction in dynamic range, tone mapping operations inevitably cause information loss, and thus a natural question is how much structural information in the HDR image is faithfully preserved in the tone mapped image. Motivated by the philosophy of the SSIM approach [10], a cross dynamic range structural fidelity measure was proposed in [14] and subsequently refined in [15, 16]. This method not only provides an overall quality assessment of a tone mapped image, but also produces a structural fidelity map that indicates how well the local structural details are preserved at each spatial location.

The purpose of this work is to develop better TMOs by making use of the objective structural fidelity measure proposed in [16]. Unlike the design of traditional TMOs, we do not start from a different computational structure for tone mapping. Instead, we explicitly treat tone mapping as an optimization problem in the image space and propose an iterative search approach that starts from any initial image and moves step-by-step in the image space towards the direction of improving the structural fidelity measure until a (local) maximal point is reached. When applied to initial images generated by existing and state-of-the-art TMOs, our algorithm almost always enhances the visibility of image details and improves the structural fidelity measure. Indeed, it often restores image structures that are missing in the images produced by state-of-the-art TMOs.

2. TONE MAPPING BY STRUCTURAL FIDELITY MAXIMIZATION

2.1. Cross Dynamic Range Structural Fidelity

Typical image similarity measures such as PSNR and SSIM [10] cannot be directly applied to assess the distortions between two im-

ages with different dynamic ranges. The cross dynamic range structural fidelity measure in [14–16] aims to overcome this problem and is briefly described as below.

Let $\mathbf{x}$ and $\mathbf{y}$ be two image patches extracted from the HDR and the LDR images, respectively. A local structural fidelity measure is defined as

$$S_{\text{local}}(\mathbf{x}, \mathbf{y}) = \frac{2\tilde{\sigma}_x\tilde{\sigma}_y + C_1}{\tilde{\sigma}_x^2 + \tilde{\sigma}_y^2 + C_1} \cdot \frac{\sigma_{xy} + C_2}{\sigma_x\sigma_y + C_2}, \quad (1)$$

where the first term is a modification of the local contrast comparison component in SSIM (more details are given later). The second term is the same as the structure comparison component in the original SSIM index, where σ_x , σ_y and σ_{xy} denote the local standard deviations and cross correlation between the two corresponding patches in the HDR and LDR images, respectively. As in SSIM, C_1 and C_2 are positive constants to avoid instabilities at low energy regions.

The local contrast comparison term in (1) is derived based on two intuitive considerations. First, as long as the contrast in the HDR and LDR patches are both significant or both insignificant, the contrast differences should not be penalized. Second, the measure should penalize the cases in which the contrast is significant in one of the patches but not the other. To gauge the significance of local contrast, we pass the local standard deviation σ through a nonlinear mapping function, resulting in the $\tilde{\sigma}$ value in (1). In [16], the nonlinear mapping function is defined using a psychometric function known as Galton's ogive [17], which determines the detection probability of the amplitude of the sinusoidal stimulus using a cumulative normal distribution function. In [16], the psychometric function was rewritten in terms of the standard deviation of the signal. As a result, the mapping between σ and $\tilde{\sigma}$ is given by

$$\tilde{\sigma} = \frac{1}{\sqrt{2\pi}\theta_\sigma} \int_{-\infty}^{\sigma} \exp\left[-\frac{(t - \tau_\sigma)^2}{2\theta_\sigma^2}\right] dt, \quad (2)$$

where τ_σ is a contrast threshold and $\theta_\sigma = \tau_\sigma/3$ [16]. In [16], τ_σ is calculated for natural images using a CSF model as well as a contrast sensitivity measurement assuming a pure sinusoidal stimulus.

The local structural fidelity measure S_{local} is applied using a sliding window that runs across the image, resulting in a map that reflects the variation of structural fidelity across space. Figure 1(g) shows an example of such a structural fidelity map computed for a tone mapped “memorial” image Fig. 1(a). The structural fidelity map is reasonably consistent with visual perception. For example, due to overexposure, the structural details of the brightest window region are missing, which are well indicated in the map. Finally, the quality map is averaged to provide a single overall structural fidelity measure of the image:

$$S(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{i=1}^M S_{\text{local}}(\mathbf{R}_i \mathbf{X}, \mathbf{R}_i \mathbf{Y}) = \frac{1}{M} \sum_{i=1}^M S_{\text{local}}(\mathbf{x}_i, \mathbf{y}_i), \quad (3)$$

where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors representing the HDR and the tone mapped LDR images, respectively, $\mathbf{R}_i$ denotes the matrix that extract the i -th patch from the image, $\mathbf{x}_i = \mathbf{R}_i \mathbf{X}$ and $\mathbf{y}_i = \mathbf{R}_i \mathbf{Y}$ are column vectors of length N representing the i -th patches extracted from the HDR and LDR images, respectively, and M is the total number of patches. Following [16], we set $C_1 = 0.01$, $C_2 = 10$, and employ a Gaussian sliding window of size 11×11 (and thus $N = 121$) with standard deviation 1.5 to create the fidelity map.

It was shown in [16] that the structural fidelity measure described above is well correlated with subjective quality evaluations of LDR images, and its performance is statistically similar to an average human subject [16].

2.2. Tone Mapping as an Optimization Problem in the Image Space

If the purpose of tone mapping is to achieve the best structural fidelity, then optimal TMO can be formulated as a maximum structural fidelity (MSF) problem given by

$$\mathbf{Y}_{\text{MSF}} = \arg \max_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y}). \quad (4)$$

This is an optimization problem in high dimension space (the same dimension as the number of pixels in the images) and finding the global optimal is difficult. Assuming smooth and regular behavior of the structural fidelity function, here we propose to use a gradient ascent algorithm to search for local optimal solutions. Similar gradient-based approaches for SSIM optimization has been studied previously for the purpose of comparing competing image quality measures but has not been explored in the context of image quality enhancement or high dynamic range imaging [18, 19].

Given an initial guess image $\mathbf{Y}_0$, we use an iterative algorithm to search along the gradient ascent direction. At the k -th iteration, the solution is updated by

$$\mathbf{Y}_k = \mathbf{Y}_{k-1} + \lambda \nabla_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y})|_{\mathbf{Y}=\mathbf{Y}_{k-1}}, \quad (5)$$

where $\nabla_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y})|_{\mathbf{Y}=\mathbf{Y}_{k-1}}$ is the gradient of $S(\mathbf{X}, \mathbf{Y})$ with respect to $\mathbf{Y}$ at the previous solution $\mathbf{Y}_{k-1}$, and λ is a constant that determines the speed of movement along the gradient direction.

To compute the gradient $\nabla_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y})$, we start from the local structural fidelity and rewrite (1) as

$$S_{\text{local}}(\mathbf{x}, \mathbf{y}) = \frac{A_1 A_2}{B_1 B_2}, \quad (6)$$

where

$$A_1 = 2\tilde{\sigma}_x\tilde{\sigma}_y + C_1 \quad (7)$$

$$B_1 = \tilde{\sigma}_x^2 + \tilde{\sigma}_y^2 + C_1 \quad (8)$$

$$A_2 = \sigma_{xy} + C_2 \quad (9)$$

$$B_2 = \sigma_x\sigma_y + C_2. \quad (10)$$

Since both image patches are represented as column vectors of length N , we have

$$\mu_y = \frac{1}{N} \mathbf{1}^T \mathbf{y} \quad (11)$$

$$\sigma_y^2 = \frac{1}{N} (\mathbf{y} - \mu_y)^T (\mathbf{y} - \mu_y) \quad (12)$$

$$\sigma_{xy} = \frac{1}{N} (\mathbf{x} - \mu_x)^T (\mathbf{y} - \mu_y). \quad (13)$$

The gradient of local structural fidelity measure with respect to $\mathbf{y}$ can then be expressed as

$$\begin{aligned} \nabla_{\mathbf{y}} S_{\text{local}}(\mathbf{x}, \mathbf{y}) &= \frac{(A'_1 A_2 + A_1 A'_2)}{B_1 B_2} \\ &\quad - \frac{(B'_1 B_2 + B_1 B'_2) A_1 A_2}{(B_1 B_2)^2}, \end{aligned} \quad (14)$$

where

$$A'_1 = \nabla_{\mathbf{y}} A_1, \quad B'_1 = \nabla_{\mathbf{y}} B_1, \quad A'_2 = \nabla_{\mathbf{y}} A_2, \quad B'_2 = \nabla_{\mathbf{y}} B_2. \quad (15)$$

Noting that

$$\nabla_{\mathbf{y}} \sigma_y = \frac{1}{N \sigma_y} (\mathbf{y} - \mu_y) \quad (16)$$

$$\nabla_{\mathbf{y}} \sigma_{xy} = \frac{1}{N} (\mathbf{x} - \mu_x), \quad (17)$$

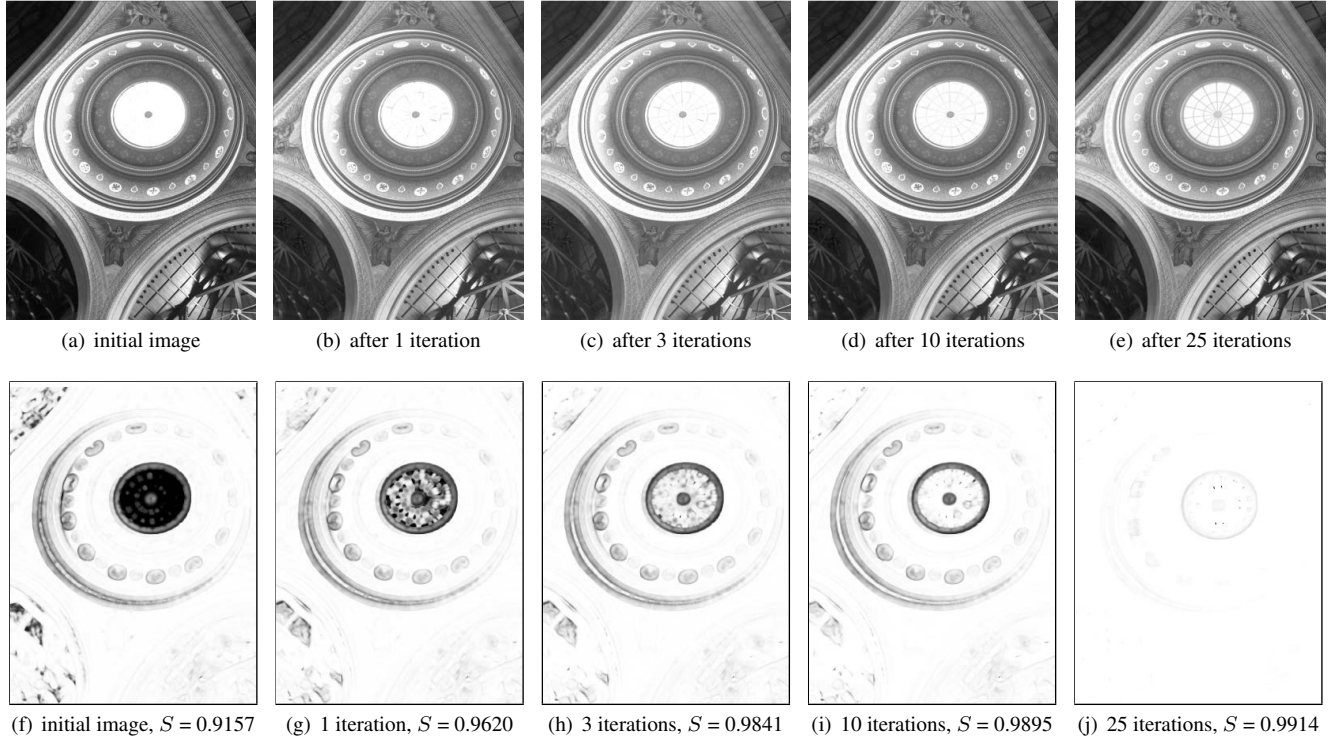


Fig. 1. Tone mapped “Memorial” images and their structural fidelity maps. The initial image (a) was created by Adobe Photoshop “Exposure and Gamma” method and (f) is its structural fidelity map, where brighter indicates higher structural fidelity. The top row also shows the images created after the first (b), the third (c), the 10-th (d) and the 25-th (e) iterations using the proposed algorithm, and the bottom row shows their corresponding structural fidelity maps (g)-(j). All images are cropped for better visualization.

we have

$$\begin{aligned} A'_1 &= 2\tilde{\sigma}_x \nabla_{\mathbf{y}} \tilde{\sigma}_y \\ &= \frac{2\tilde{\sigma}_x}{\sqrt{2\pi}\theta_\sigma} \exp\left[-\frac{(\sigma_y - \tau_\sigma)^2}{2\theta_\sigma^2}\right] \cdot \nabla_{\mathbf{y}} \sigma_y \\ &= \sqrt{\frac{2}{\pi}} \frac{\tilde{\sigma}_x}{N\theta_\sigma \sigma_y} \exp\left[-\frac{(\sigma_y - \tau_\sigma)^2}{2\theta_\sigma^2}\right] (\mathbf{y} - \mu_y), \end{aligned} \quad (18)$$

$$\begin{aligned} B'_1 &= 2\tilde{\sigma}_y \nabla_{\mathbf{y}} \tilde{\sigma}_y \\ &= \sqrt{\frac{2}{\pi}} \frac{\tilde{\sigma}_y}{N\theta_\sigma \sigma_y} \exp\left[-\frac{(\sigma_y - \tau_\sigma)^2}{2\theta_\sigma^2}\right] (\mathbf{y} - \mu_y), \end{aligned} \quad (19)$$

$$A'_2 = \frac{1}{N} (\mathbf{x} - \mu_x), \quad (20)$$

$$B'_2 = \sigma_x \nabla_{\mathbf{y}} \sigma_y = \frac{\sigma_x}{N\sigma_y} (\mathbf{y} - \mu_y). \quad (21)$$

Plugging (7), (8), (9), (10), (18), (19), (20) and (21) into (14), we obtain the gradient of local structural fidelity. Finally, summing all the local gradient together, we can compute the the gradient of the overall structural fidelity measure with respect to the LDR image $\mathbf{Y}$ as

$$\nabla_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{i=1}^M \mathbf{R}_i^T \nabla_{\mathbf{y}} S_{\text{local}}(\mathbf{x}, \mathbf{y})|_{\mathbf{x}=\mathbf{x}_i, \mathbf{y}=\mathbf{y}_i}, \quad (22)$$

which is subsequently plugged into (5) to update the solution that is fed into the next iteration.

3. EXPERIMENTAL RESULTS

There are only two new parameters that need to be determined in our iterative algorithm, which are λ that controls the speed of convergence, and the termination threshold T that stops the iteration when the structural fidelity values between two consecutive iterations is smaller than the threshold. Throughout our experiment, we set $\lambda = 150$ and $T = 0.0001$ and empirically find the behavior of the proposed algorithm is not sensitive to these parameters.

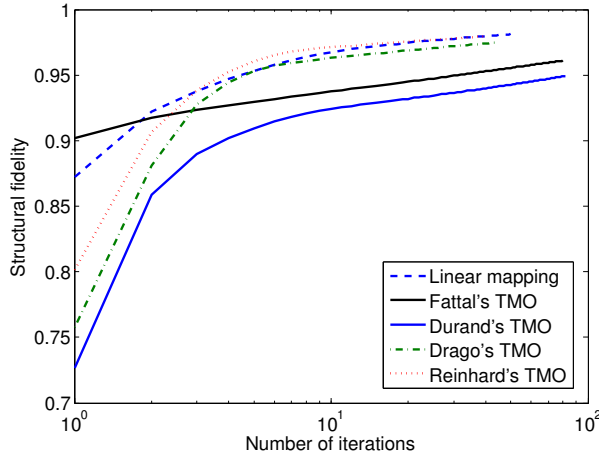
To test the performance of the proposed method, we select a set of widely-used HDR images as test images, which are partially listed in Table 1 and similar results are obtained for other test images. The initial images for the iterative algorithm are created using different approaches, including blank images (all image pixels are set to 128), linearly mapped images (direct linear scaling from the source HDR images to the dynamic range of $[0, 255]$), and images created by state-of-the-art TMOs [2, 4, 5, 20]. It can be observed in Table 1 that starting from simple blank and linearly-scaled initial images, the proposed method successfully produces high structural fidelity images. In addition, it is also quite effective at improving upon all state-of-the-art TMOs.

Figure 1 provides a visual demonstration of the iterative procedure, where the proposed algorithm is applied to an initial tone mapped “Memorial” image created by the “Exposure and Gamma” method in Adobe Photoshop. The structural fidelity map is very effective at detecting the missing details in the tone mapped images. For example, the structural details in the brightest window region in the initial image Fig. 1(a) are completely lost due to tone mapping

Table 1. Comparison of structural fidelity scores between initial and converged images

Image		Blank image	Linear mapping	Fattal [4]	Durand [20]	Reinhard [2]	Drago [5]
Kitchen	initial image	0.0099	0.7367	0.8748	0.7467	0.8073	0.7728
	converged image	0.8757	0.9737	0.9721	0.9758	0.9840	0.9796
Bristol	initial image	0.0099	0.8726	0.9020	0.7269	0.8013	0.7574
	converged image	0.8754	0.9812	0.9610	0.9493	0.9796	0.9747
Tinterna	initial image	0.0099	0.9572	0.9471	0.8797	0.9473	0.9307
	converged image	0.9052	0.9951	0.9786	0.9818	0.9958	0.9949
Memorial	initial image	0.0099	0.9573	0.9586	0.9559	0.9832	0.9765
	converged image	0.9184	0.9934	0.9863	0.9950	0.9959	0.9972

and are clearly indicated by the central dark region in the structural fidelity map Fig. 1(f). With the progress of iterations, such details are becoming more and more visible until nearly perfectly restored after 25 iterations as shown in Fig. 1(e). The evolvement of the structural details is very well tracked by the structural fidelity maps computed along with the iterations, which eventually converge to a nearly uniform white picture.

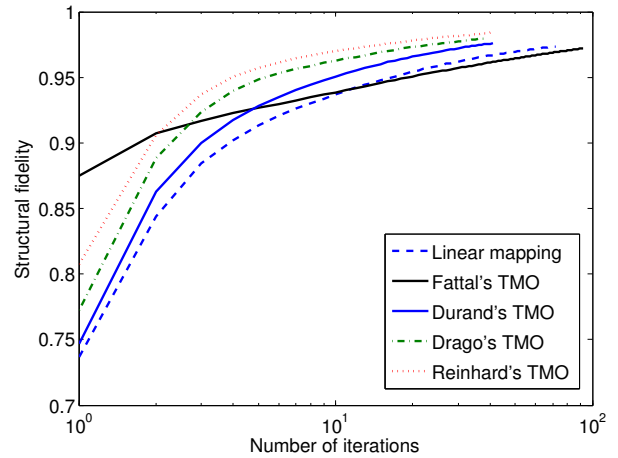
**Fig. 2.** Structural fidelity versus iteration of tone mapped “Bristol” images with initial images created by different TMOs.

To observe the behavior of the iterations numerically, we plot the structural fidelity measure as a function of iteration for two source images in Figs. 2 and 3, respectively. It appears that the proposed iterative approach is well-behaved and always increases monotonically until it converges to a fixed point in the image space, although the fixed point may only be a local maximum. The different structural fidelity values of the converged images suggest that the local maxima obtained from different initial images are different. This indirectly reflects the complication of the search space.

The computational complexity in each iteration of the proposed algorithm is linear with respect to the number of pixels in the image. Our unoptimized MATLAB implementation on an Intel Quad-Core 2.67GHz computer takes on average around 6 seconds per iteration on an image of size 512×512 .

4. CONCLUSIONS AND DISCUSSIONS

In this study, we formulate tone mapping as an optimization problem in the space of images and propose a gradient ascent algorithm to

**Fig. 3.** Structural fidelity versus iteration of tone mapped “Kitchen” images with initial images created by different TMOs.

maximize the quality of the tone mapped image in terms of a recently proposed structural fidelity measure. Starting from simple images such as a completely blank image or a linearly scaled image from the HDR image, the proposed algorithm can automatically find images with reasonably high structural fidelity. When it is applied to initial images created by state-of-the-art TMOs, it always produces even better quality images.

The monotonic convergence property and the analysis of Figs. 2 and 3 suggest a *locally smooth/convex* and *globally sophisticated* high-dimensional search space. The limitation of the current gradient ascent algorithm is that it can only find local maxima. Improved solutions may be found by starting from multiple initial images and pick the resulting converged image that has the highest structural fidelity. However, this still cannot guarantee global maximum. In the future, the search space needs to be carefully investigated and more advanced optimization algorithms can be exploited. In addition, other quality factors such as natural scene statistics models [15, 16, 21] may also be incorporated to further improve the results.

5. REFERENCES

- [1] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann Publishers Inc., 2010.
- [2] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *Proc. of 29th annual Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, vol. 21, pp. 267–276, 2002.
- [3] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 4, pp. 291–306, 1997.
- [4] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '02*, pp. 249–256, 2002.
- [5] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum*, vol. 22, no. 3, pp. 419–426, 2003.
- [6] F. Drago, W. L. Martens, K. Myszkowski, and S. H. P., "Perceptual evaluation of tone mapping operators," In *Proc. Of the SIGGRAPH Conf. Sketches and Applications*, 2003.
- [7] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, "Image attributes and quality for evaluation of tone mapping operators," in *Proceedings of the 14th Pacific Conference on Computer Graphics and Applications*, (Taipei, Taiwan), pp. 35–44, National Taiwan University Press, 2006.
- [8] C. Cavaro-Menard, L. Zhang, and P. Le Callet, "Diagnostic quality assessment of medical images: Challenges and trends," in *2010 2nd European Workshop on Visual Information Processing, EUVIP2010*, pp. 277–284, 2010.
- [9] Z. Wang, "Applications of objective image quality assessment methods," *IEEE Signal Processing Magazine*, vol. 28, pp. 137–142, Nov. 2011.
- [10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, pp. 600–612, Apr. 2004.
- [11] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, pp. 98–117, Jan. 2009.
- [12] R. Mantiuk, S. Daly, K. Myszkowski, and Seidel, "Predicting visible differences in high dynamic range images - model and its calibration," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5666, pp. 204–214, 2005.
- [13] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H. . Seidel, "Dynamic range independent image quality assessment," in *SIGGRAPH'08: International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, 2008.
- [14] H. Yeganeh and Z. Wang, "Objective assessment of tone mapping algorithms," in *Proc. IEEE Int. Conf. Image Proc.*, 26–29 Sept., China 2010.
- [15] H. Yeganeh and Z. Wang, "Structural fidelity vs. naturalness - objective assessment of tone mapped images," (*LNCS*) *International Conference on Image Analysis and Recognition, Burnaby, BC, Canada*, vol. 6753, June 2011.
- [16] H. Yeganeh and Z. Wang, "Objective quality assessment of tone mapped images," *IEEE Trans. on Image Processing*, vol. 22, pp. 657–667, Feb. 2013.
- [17] P. G. J. Barten, *Contrast sensitivity of the human eye and its effects on image quality*. Washington: SPIE Optical Engineering Press, 1999.
- [18] Z. Wang and E. P. Simoncelli, "Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics," in *Human Vision and Electronic Imaging IX, Proc. SPIE*, vol. 5292, pp. 99–108, Jan. 2004.
- [19] Z. Wang and E. P. Simoncelli, "Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities," *Journal of Vision*, vol. 8, pp. 1–13, Sept. 2008.
- [20] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *ACM Transactions on Graphics*, vol. 21, pp. 257–266, 2002.
- [21] Z. Wang and A. C. Bovik, "Reduced- and no-reference visual quality assessment - the natural scene statistic model approach," *IEEE Signal Processing Magazine*, vol. 28, pp. 29–40, Nov. 2011.